REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE 11/17/03	3. REPORT TYPE AND DATES COVERED Final, 6/1/00 – 11/30/01

4. TITLE AND SUBTITLE
Hidden Markov Models for Sensor Fusion of EMI and GPR

5. FUNDING NUMBERS
DAAD19-00-1-0372

6. AUTHOR(S)
Paul Gader

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Missouri – Columbia
201 EBW
Columbia, MO 65211

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

41339.1-EV

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT	12 b. DISTRIBUTION CODE
Approved for public release; distribution unlimited.	

13. ABSTRACT (Maximum 200 words)

An investigation of hidden Markov models (HMMs) as a means of processing Ground Penetrating Radar signals for discrimination between anti-tank (AT) and anti-personnel (AP) landmines and discrete clutter objects. Hidden Markov models had been used successfully previously in AT landmine detection with GPR but had not been used as a discriminant between mines and clutter nor had they been used on AP mines. Experiments were conducted on data collected at the JUXOCO calibration grid. Continuous and discrete HMMs were trained and tested and evaluated on the grid using both the Baum-Welch and discriminative training algorithms. Experimental results suggest that discriminative training algorithms should be used for training HMMs for landmine detection, fusion of continuous and discrete models provided improved performance, and that, although HMMs with the existing feature sets perform well for detection of AT mines, new feature sets should be developed for discriminating between mines and clutter objects.

14. SUBJECT TERMS Landmine detection, ground penetrating radar, hidden Markov models, discriminative training			15. NUMBER OF PAGES 20
			16. PRICE CODE
17. SECURITY CLASSIFICATION OR REPORT **UNCLASSIFIED**	18. SECURITY CLASSIFICATION ON THIS PAGE **UNCLASSIFIED**	19. SECURITY CLASSIFICATION OF ABSTRACT **UNCLASSIFIED**	20. LIMITATION OF ABSTRACT **UL**

Final Technical Report

ARO Research Agreement No. DAAD19-00-1-0372

Paul Gader, Principal Investigator

Statement of problem studied. An investigation of hidden Markov models (HMMs) as a means of processing Ground Penetrating Radar signals for discrimination between anti-tank (AT) and anti-personnel (AP) landmines and discrete clutter objects was conducted. Hidden Markov models had been used successfully previously in AT landmine detection with GPR but had not been used as a discriminant between mines and clutter nor had they been used on AP mines. Experiments were conducted on data collected at the JUXOCO calibration grid.

Summary of the most important results. Continuous and discrete HMMs were trained and tested and evaluated on the grid using both the Baum-Welch and discriminative training algorithms. Experimental results suggest that discriminative training algorithms should be used for training HMMs for landmine detection, fusion of continuous and discrete models provided improved performance, and that, although HMMs with the existing feature sets perform well for detection of AT mines, new feature sets should be developed for discriminating between mines and clutter objects.

Table of Contents

List of Illustrations and Tables

List of participating scientific personnel. Ms. Wei Xiong was a graduate research assistant on the project at the University of Missouri – Columbia. She has almost completed her M. S. degree on this topic. Dr. Hong-Chi Shi of the faculty of the University of Missouri – Columbia also worked on the project.

List of publications and technical reports. Reference [1] in the bibliography below was impacted by this research.

1. GPR System

GPR data from the GEO-CENTERS Energy Focusing Ground Penetrating Radar (EFGPR) system were used in these experiments. The EFGPR is a time-domain impulse radar system, which includes an array of antennas, a synchronous high-speed interface model, a GPS system, and a host computer for control and processing. It uses delayed signals in a wide multi-element array which focuses transmitted and received signals to locate targets in the soil.

GEO-CENTERS developed several models of the EFGPR. The model used in this project is Model 401 EFGPR. Model 401 EFGPR is a portable, three-wheeled humanitarian de-mining system. It deploys Rolled Edge Transverse Electromagnetic (RETEM) antennas that increase in both gain and upper bandwidth, along with improvements in RF components to take advantage of the enhanced radar bandwidth.. It supports a single 1.5m, 6-antenna-pair-array to cover a 1m-detection swath. The center frequency of the EFGPR is 1.25GHZ. The 401 EFGPR saves the GPR to disk for off-line analysis.

A raster scan is generated every 5cm when the system advances. The scan is stored in 25 x 40 x 12 bit radar image format that represents a ground slice. The 25 pixels cross track represent 1m wide swath, and the 40 pixels represent the time or depth. Figure 1 is an illustration of a landmine signature. The down-track direction is the direction of vehicle motion. The cross-track direction is represented by the variable x, the down-track direction by the variable y, and the depth by the variable z.

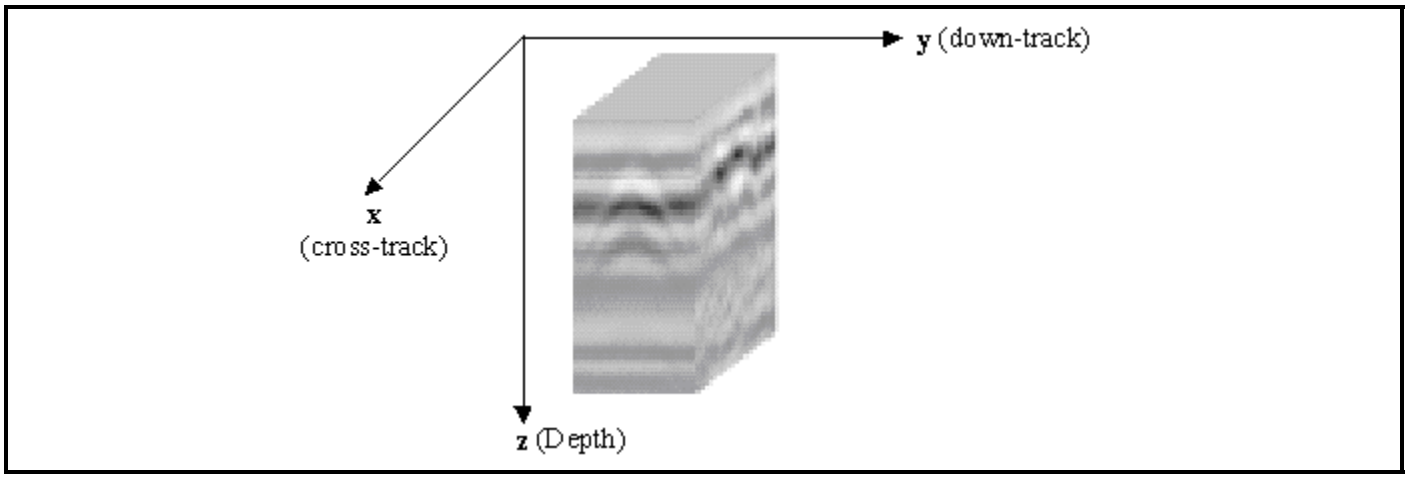

Figure 1: An Illustration of Three-dimensional Landmine Signature

2. JUXOCO Site

The GEO-CENTERs' GPR collected data at a JUXOCO calibration grid in May 2001. The grid has since been dismantled. The calibration grid consisted of a 5m x 25m plot. The grid was divided into 5 lanes side by side, referred to as lanes A, B, C, D, and E. Each lane was 25m long and 1m wide. Each lane was uniformly divided into 25 grid squares. At the center of each square, a landmine or clutter was buried in various depths, or the square was left empty. In lane A and lane C, different types of clutter were buried at varies depths in the grid squares. The anti-personnel landmines were buried in some grid squares of Lane B, while anti-tank landmines in lane D. In lane E, AP and AT landmines, landmine simulants and several clutter objects were buried in most of the squares. Table 1 is an enumeration of the objects buried in the grid. The weights of the metal clutter objects range from 0.41g to 306.00g. The non-metal clutter objects include woods, stones, a plastic spray paint cap, and a filled hole.

Landmines/Landmine Simulants			Objects	
AT Landmines	AP Landmines	Landmine Simulants	Non-Metal Objects	Metal Objects
17	10	2	17	16
29			33	

Table 1: Objects buried in the calibration grid. The other grid squares contained nothing.

3. Hidden Markov Models

3.1. HMM Description

Hidden Markov models are stochastic models for stochastic processes that produce time sequences of random observations as a function of states. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state, an output or observation can be generated according to the associated probability distribution. It is only the output, not the state, that is visible to an external observer. Therefore states are hidden or not observable to the outside. Although HMMs are described elsewhere, e.g. [1-3], we provide some description here for completeness.

The input to the HMM is the observation sequence, which is a sequence of feature vectors $O = o_1 o_2 \cdots o_T$. The number of states of the model is given by *N,* the individual states by

$S = \{S_1, S_2, \cdots, S_N\}$, and the current state by q_t. For a discrete HMM, the number of distinct observation symbols is *M*, and the individual symbols are given by $V = \{v_1, v_2, \cdots, v_m\}$. Every HMM has a set of state transition probabilities $A = \{a_{ij}\}$ given by

$$a_{ij} = p\{q_{t+1} = S_j \mid q_t = S_i\},\ 1 \le i, j \le N \tag{1}$$

Transition probabilities should satisfy the normal stochastic constraints,

$$a_{ij} \ge 0 \text{ and } \sum a_{ij} = 1,\ 1 \le i, j \le N \tag{2}$$

A discrete HMM is has a set of discrete, conditional probability density functions, one for each state,which can be used to form the matrix , $B = \{b_j(k)\}$ where

$$b_j(k) = p\{o_t = v_k \mid q_t = S_j\}, 1 \le j \le N,\ 1 \le k \le M \tag{3}$$

is the observation sequence. The following stochastic constraints must be satisfied

$$b_j(k) \ge 0 \text{ and } \sum_{k=1}^{M} b_j(k) = 1,\ 1 \le j \le N,\ 1 \le k \le M. \tag{4}$$

A continuous HMM uses continuous probability density functions. In this case we specify the parameters of the probability density function. Usually the probability density is approximated by a weighted sum of *M* Gaussian distributions N,

$$b_j(o_t) = \sum_{m=1}^{M} c_{jm} N(\mu_{jm}, \Sigma_{jm}, o_t), \tag{5}$$

Where the c_{jm} are the mixture coefficients, μ_{jm} the mean vectors, and $\sum_{jm}$ thecovariance matrices. The coefficients c_{jm} must satisfy the stochastic constraints

$$c_{jm} \ge 0 \text{ and } \sum_{m=1}^{M} c_{jm} = 1, 1 \le j \le N,\ 1 \le k \le M. \tag{6}$$

The initial state distributions are given by $\pi = \{\pi_i\}$, where $\pi_i = p\{q_1 = i\}$, $1 \le i \le N$.

Therefore we can use the compact notation $\lambda = (A, B, \pi)$ to denote an HMM with discrete probability distributions, and $\lambda = (A, c_{jm}, \mu_{jm}, \Sigma_{jm}, \pi)$ to denote one with continuous densities.

The input to the HMM is the observation sequence, which is a sequence of feature vectors $O = o_1 o_2 \cdots o_T$. The output of an HMM is computed using the Viterbi algorithm to find the optimal state sequence and produces the quantity $\log\left(P\left(O, q^* \mid \lambda\right)\right)$ where q^* represents the optimal state sequence [1]. This quantity can be thought of as representing the probability of the observation sequence given the model.

In the landmine detection problem, observation vectors are generated from GPR measurements as completely described in [2]. The states are associated with varying geometry between the GPR antennas and the object, as depicted in Figure 2. Initially the system may be in a background state, then changes to a state in which the sensor receives returns from the mines but is not over the mine, the one in which it is over the mine, then moving away from the mine but still receiving returns from the mine, and finally in a background state again.

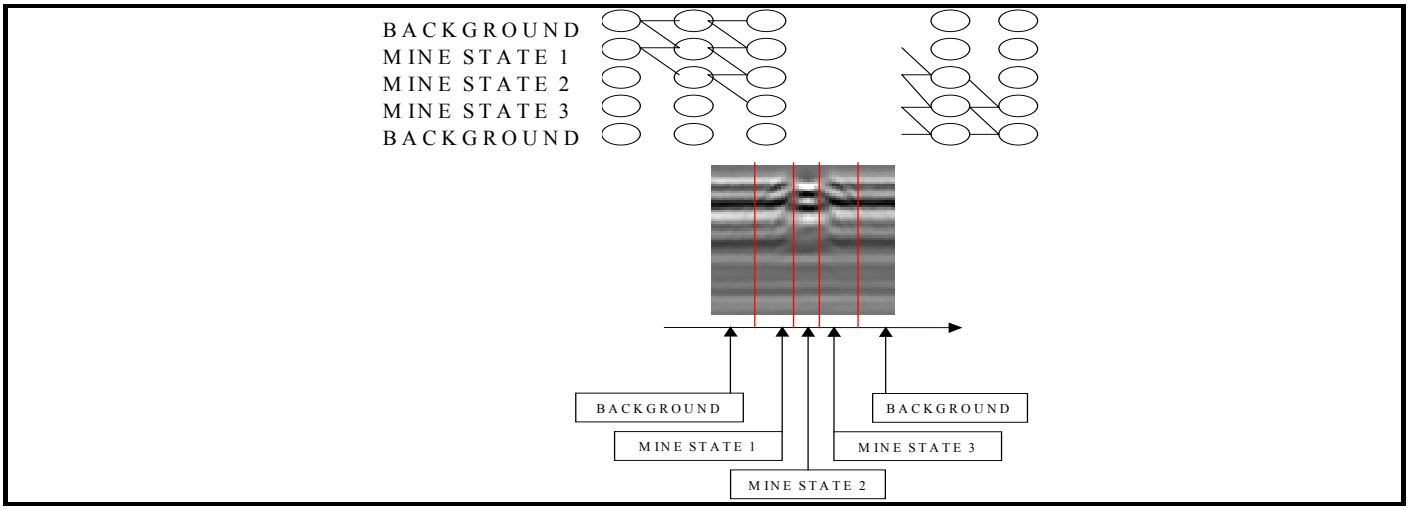

Figure 2. Depiction of a mine HMM for landmine detection.

There is one observation vector per channel per measurement. Sequences of 15 observation vectors, corresponding to approximately 0.75 meter are used as inputs to the HMM. For a given experiment, there are two models: a mine model and a background model. The output of the HMM landmine detection system is a confidence value given by

$$C = \log\left(P\left(O, q^* \mid \lambda_{\text{mine}}\right)\right) - \log\left(P\left(O, q^* \mid \lambda_{background}\right)\right) = \log\left(P\left(O, q^* \mid \lambda_{\text{mine}}\right) \Big/ P\left(O, q^* \mid \lambda_{background}\right)\right) \qquad (7)$$

3.2. Training

Training an HMM is the process of estimating the parameters of the HMM using a training set. The standard method of training HMMs is the Baum-Welch method, which is well-described in the literature. In addition to Baum-Welch, it has been shown that Minimum Classification Error (MCE) training, also referred to as discriminative training, improves performance of HMMs for landmine detection [3]. These methods are described in the literature but are included here for completeness. In addition, the MCE training for landmine detection discussed in the literature is for discrete HMMs only. The results shown in this document are the first reported results using MCE training with continuous HMMs for landmine detection.

3.2.1. Baum-Welch Training

To solve the learning problem, Baum and his colleagues defined an auxiliary function [4]:

$$Q(\lambda,\overline{\lambda}) = \sum_{q} p(q \mid O,\lambda)\log\left(p(O,q \mid \overline{\lambda})\right) \tag{8}$$

where $\overline{\lambda}$ is the auxiliary variable corresponding to λ. They also proved if the value of $Q(\lambda,\overline{\lambda})$ increases, then the value of $p(O \mid \overline{\lambda})$ also increases, i.e.

$$Q(\lambda,\overline{\lambda}) \geq Q(\lambda,\lambda) \rightarrow p(O \mid \overline{\lambda}) \geq p(O \mid \lambda) \tag{9}$$

Thus, the problem of maximising $p(O \mid \lambda)$ is replaced by maximizing $Q(\lambda,\overline{\lambda})$ with respect to $\overline{\lambda}$. Two auxiliary variables are defined for use in Baum-Welch training. The first variable is defined as the probability of the t^{th} observation being in state S_i and the $(t+1)^{\text{th}}$ observation being in state S_j. Formally,

$$\xi_t(i,j) = p(q_t = S_i, q_{t+1} = S_j \mid O,\lambda) \tag{10}$$

$$\xi_t(i,j) = \frac{p(q_t = S_i, q_{t+1} = S_j, O \mid \lambda)}{p(O \mid \lambda)} \tag{11}$$

Using forward and backward variables this can be expressed as,

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} \beta_{t+1}(j) b_j(o_{t+1})}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i) a_{ij} \beta_{t+1}(j) b_j(o_{t+1})} \tag{12}$$

The second variable is a posteriori probability, which is the probability of the t^{th} observation being in state S_i:

$$\gamma_t(i) = p\left(q_t = S_i \mid O, \lambda\right) \tag{13}$$

Using forward and backward variables we can rewrite equation (11) as,

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{14}$$

One can see that the relationship between $\gamma_t(i)$ and $\xi_t(i,j)$ is given by,

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j), \; 1 \le i \le N, \; 1 \le t \le M \tag{15}$$

In the Baum-Welch learning process, the parameters of a discrete HMM are updated in such a way to maximize $p(O \mid \lambda)$. Assuming a starting model $\lambda = (A, B, \pi)$, one calculates the following so-called re-estimation formulas

$$\overline{\pi_i} = \gamma_1(i), \qquad 1 \le i \le N, \tag{16}$$

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i,j)}, \qquad 1 \le i, j \le N, \tag{17}$$

$$\overline{b_j}(k) = \frac{\sum_{t=1, O_t = \upsilon_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}, \qquad 1 \le j \le N, \; 1 \le k \le M \tag{18}$$

Reestimation formulas have been derived for to the continuous density case [5][6].

3.2.2. Minimum Classification, or Discriminative, Training

Minimum Classification Error (MCE) with Generalized Probabilistic Descent was first proposed by Juang and Katagiri [7] based on an earlier approach by Amari[8]. It has been widely applied to several

classifier structures, such as Multi-Layer Perceptrons [9], and Hidden Markov Models [7]. The essential aspect of this approach is to train the classifier structure so as to minimize the classification error rate using a gradient-based method together. However the error rate involves a discontinuous classification loss, since the classification is either correct or incorrect. This makes it difficult to apply gradient-based optimization techniques, which requires that the objective loss function is at least first-order differentiable. The strategy of MCE is to smooth the discontinuous classification loss function, while still staying close to the loss function and use a gradient-based adaptation method.

Consider a set of observations $L = \{x_1, x_2, x_3, \cdots, x_N\}$, where x_i is from one of the M classes C_j, $j = 1, 2, \cdots, M$, a classifier parameter set Λ, M discriminant functions $g_j(x;\Lambda)$, and the decision rule:

$$C(x) = C(i), \text{ if } g_i(x;\Lambda) = \max_j g_j(x;\Lambda) \tag{19}$$

A general misclassification measure for the k^{th} class sample can be defined as [7]:

$$d_i(x) = -g_i(x;\Lambda) + \left[\frac{1}{M-1} \sum_{j \neq i} g_j(x;\Lambda)^{\eta} \right]^{\frac{1}{\eta}}, \tag{20}$$

where η is a positive number, controls the contribution of each misclassification towards the error metric. Note that when η is large, the most confusable class contributes the most to the summation component. The cost function can be defined as [7]:

$$\ell_k(x;\Lambda) = \ell_k(d_k(x)), \tag{21}$$

There are many choices for the loss $\ell_k(d_k)$, for instance sigmoid function:

$$\ell_d(d_k) = \frac{1}{1 + e^{-\gamma d_k + \theta}} \ , \ \gamma > 0 \tag{22}$$

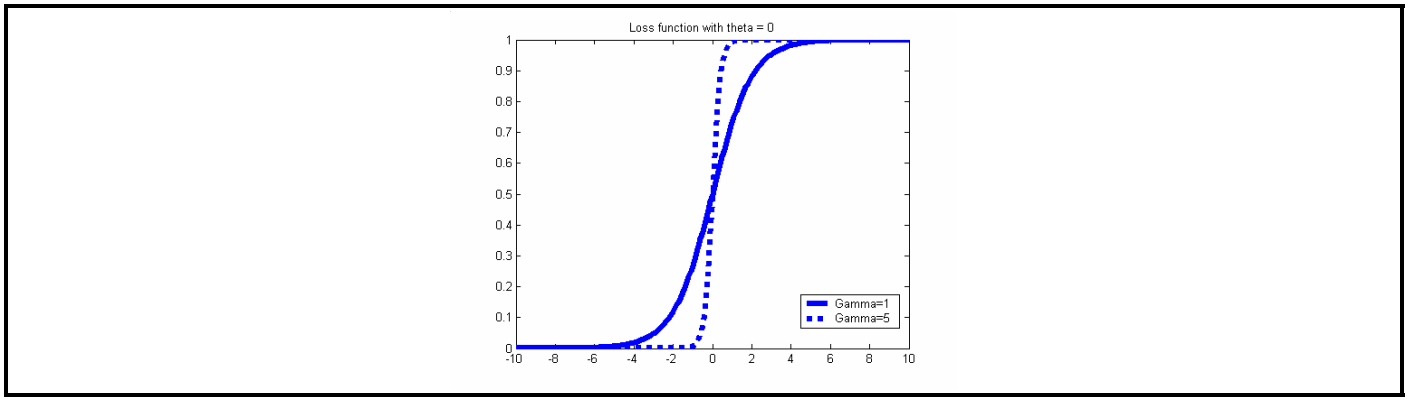

Figure 3: Sigmoid Loss Function

One can see that this loss function approximates an ideal binary loss function well and is continuous, which is suitable for gradient algorithms. A negative $d_k(x)$ indicates correct classification, and no loss is incurred; a positive $d_k(x)$ leads to a loss, which can be used to count classification error. Thus for a given sample x, the overall expected loss $L(\Lambda)$ is [7]:

$$L(\Lambda) = \sum_k p(C_k) \int l_k(x, \Lambda) p(x \mid C_k) dx \ , \tag{23}$$

Where $p(C_k)$ and $p(x \mid C_k)$ are the class a priori and conditional probabilities respectively.

Generalized Probabilistic Descent (GPD) is used to train HMMs to minimize the expected classification error, that is, to minimize the overall expectation of theloss $L(\Lambda)$. Since the distributions are unknown, the expected loss is not known. A solution to this difficulty is given by Amari's Probabilistic Descent Theorem [8] which shows that for an infinite sequence of random samples $\{x_t\}_{t=1}^{\infty}$ and step size sequence ε_t that satisfies the conditions

$$i) \sum_{t=1}^{\infty} \varepsilon_t \to \infty, \tag{24}$$

and

$$ii) \sum_{t=1}^{\infty} \varepsilon_t^2 < \infty, \tag{25}$$

adapting the system parameters according to

$$\Lambda_{t+1} = \Lambda_t - \varepsilon_t U \nabla \ell_k(x_t, \Lambda_t), \tag{26}$$

where Λ_t denotes the parameters of Λ at t, converges with a probability of one to a local minimum of $L(\Lambda)$. U is a positive definite matrix which allows one to scale the learning rate differently for different model parameters. This is important when models are more sensitive to some parameters than others, as is the case for continuous HMMs. Continuous HMMS are much more sensitive to the covariance parameters than the means and therefore different learning rates should be used . The simplest example for U is the identity matrix.

Equation 26 is used to update the parameters of the HMMs for discriminative training. Specifically, the discriminant function are

$$g_M(\mathbf{O},\lambda_M) = \log b^M_{q_0^M}(\mathbf{o}_0) + \sum_{t=1}^{T} \log a^M_{q_{t-1}^M q_t^M} + \sum_{t=1}^{T} \log b^M_{q_t^M}(\mathbf{o}_{q_t^M})$$

for the mine model (the background model is similar). In this expression, $\{ q_0^M, q_1^M, \cdots, q_T^M \}$ is the optimal state sequence found by the Viterbi algorithm.

If **O** is a misclassified background sequence, the misclassification measure is:

$$d_B(\mathbf{O}) = g_M(\mathbf{O},\lambda_M) - g_M(\mathbf{O},\lambda_M) \tag{27}$$

and the related loss function is:

$$\ell_B(d_B(\mathbf{O})) = \frac{1}{1+e^{-\gamma_B * d_B(\mathbf{O}) + \theta_B}}. \tag{28}$$

If **O** is a misclassified mine sequence, the misclassification measure is:

$$d_M(\mathbf{O}) = g_B(\mathbf{O},\lambda_B) - g_M(\mathbf{O},\lambda_M), \tag{29}$$

and the related loss function is:

$$\ell_M(d_M(\mathbf{O})) = \frac{1}{1+e^{-\gamma_M * d_M(\mathbf{O}) + \theta_M}}. \tag{30}$$

The overall loss function is:

$$\mathcal{L}(\mathbf{O})=\begin{cases}\ell_M & if \quad \mathbf{O}\in\lambda_M \\ \\ \ell_B & if \quad \mathbf{O}\in\lambda_B\end{cases}. \tag{31}$$

In order to maintain the original HMM constraints, such as:

$$\sum_{j=1}^{N} a_{ij}=1 \ \forall i=1, N, \text{ and } \sum_{i=1}^{N}\pi_i=1, \tag{32}$$

the following transformations are introduced:

$$a_{ij}=\frac{\tilde{a}_{ij}}{\sum_{j=1}^{N}\tilde{a}_{ij}} \quad \text{and} \quad \pi_i=\frac{\tilde{\pi}_i}{\sum_{j=1}^{N}\tilde{\pi}_i}. \tag{33}$$

The mine and background models are then updated according to equation (26)

4. Experimental Results

4.1. Experiments

The EFGPR made four passes over the calibration grid with different gain settings. We used two strategies for training and testing. The first strategy involved training on one pass over the calibration grid and testing on another. The second strategy involved a modified leave-one-out method, in which a given pass was iteratively divided into training and testing sets by leaving one sample out of the training set on each iteration. Each sample is left out once and only once. The feature sequences from each grid square were treat as one sample. Thus there are 125 samples in one data set. For each iteration, the initial models were trained on the training subset. The trained models were tested on the testing subset. The trained models serve as the initial models for the next run. The procedure would run for 125 times, until each sample was tested once and only once. Finally, these models were tested on the other 3 passes.

The first strategy has the advantage of different looks at the calibration lane and the disadvantage that different hardware settings were used to collect the data. The second strategy has the advantage of using the same hardware settings and that each grid square was used as a test sample exactly once and was independent of the training data. The first strategy will be referred to as the train one/test three strategy

and the second as the leave-one-out strategy. Discrete and continuous HMMs were investigated as was the fusion of discrete and continuous HMMs.

As noted above, the EFGPR collects a three-dimensional array of data corresponding to a physical region approximately one meter wide and the length of an entire lane. The JUXOCO grid was designed for sensors to collect data only at the center of the grid. Thus, the HMMs were run only on GPR data collected only near the center of each grid square. More specifically, the HMMs were run on channel thirteen of the twenty five channels and at the three down-track positions determined to be closest to the center of the grid square. Thus, three confidence values were produced for each grid square. The average of the three values was used as the confidence that a mine was present in a given grid square.

The mine and the background models were applied at each observation vector O_i^k, $i = 1, 2, 3$, where k denotes the k^{th} grid square in the lane. Each model produced an output value for each observation vector. A confidence value was produced based on the difference of the output values for each observation vector:

$$C_i^k = \log\left(P\left(O_i^k, q \mid \lambda_{\min e}\right)\right) - \log\left(P\left(O_i^k, q \mid \lambda_{background}\right)\right), \tag{34}$$

Where $i = 1, 2, 3,\ k = 1, 2, \ldots, 25.$

The average of the three confidence values was associated with the grid square.

$$C^k = \underset{i=1,2,3}{avg}\, C_i^k,\ k = 1, 2, \ldots, 25. \tag{35}$$

Further, the values of C^k were thresholded at various values to make the final decision.

$$\begin{aligned} D^k &= 1, \quad \text{if } C^k > t, \\ &= 0, \quad \text{else.} \end{aligned} \tag{36}$$

Algorithm evaluation is carried out using ROC curves. A mine is detected, if D^k is 1 and there is a mine or mine simulant present in the grid square. A mine is missed, if D^k is 0 and there is a mine or mine simulant in the grid square. A false alarm is generated, if D^k is 1 and there is no mine or mine simulant in the grid square. A background is detected, if D^k is 0 and there is no mine or mine simulant in the grid square. The probability of detection (PD) and the probability of false alarm (PFA) are computed for each value of the threshold, resulting in a ROC curve. PD is defined as the number of mines detected divided by the total number of mines and mine simulants in the grid, while PFA as the number of false alarms divided by the total number of grid squares without mine.

4.2. Results

4.2.1. Continuous HMM Results

The train one/test three strategy was used to perform experiments with continuous HMMs. Continuous HMM models used in previous AT mine detection projects were used as the initial models. These models were trained on one pass of data and evaluated on the other three passes. A total of 87 mine observation vectors and 288 background/clutter observations were used for training. The number of Gaussian mixture components in the models were varied. Continuous HMM results are shown in Figures 4-6. Since the models were tested on three different passes, the standard deviation in the Pd for each Pfa is shown also.

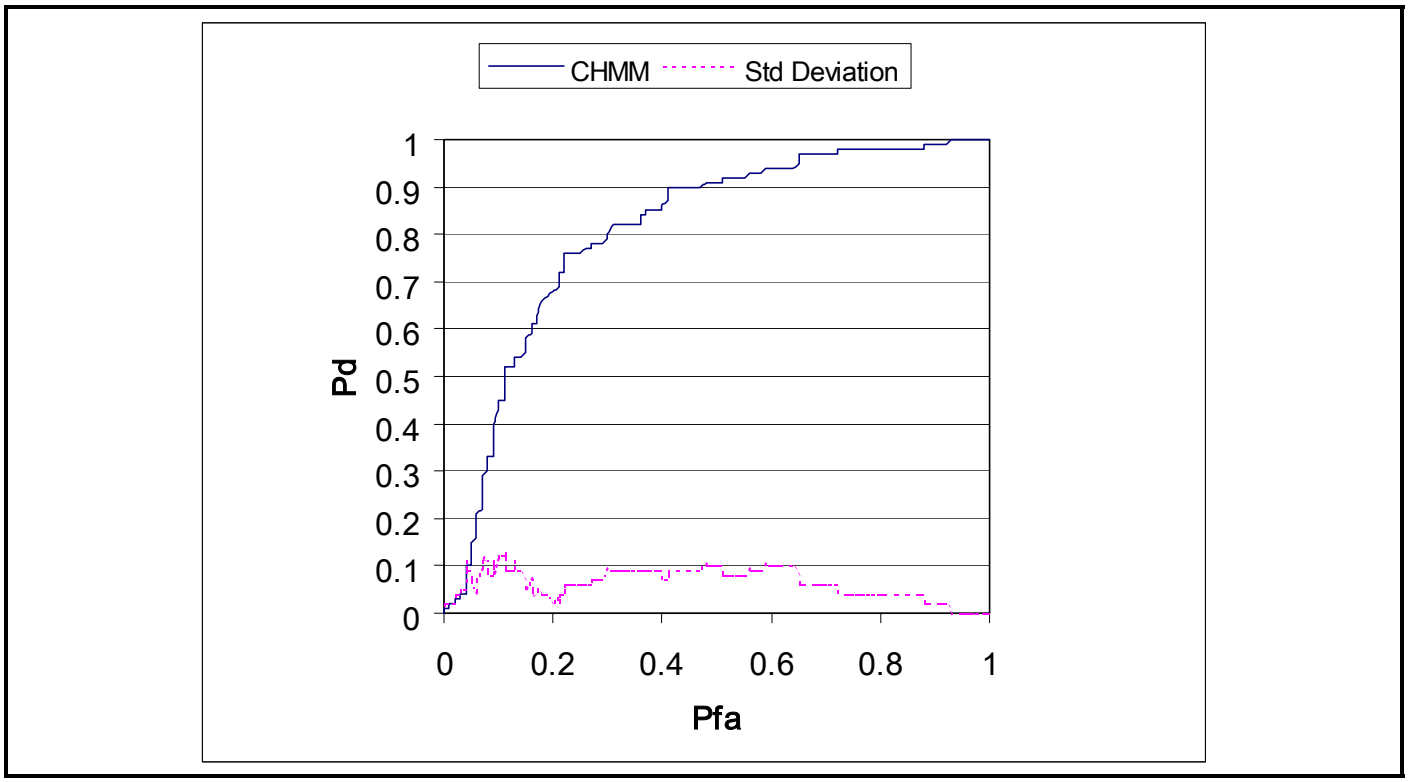

Figure 4: Results of Continuous HMM with 3 Gaussian Mixtures

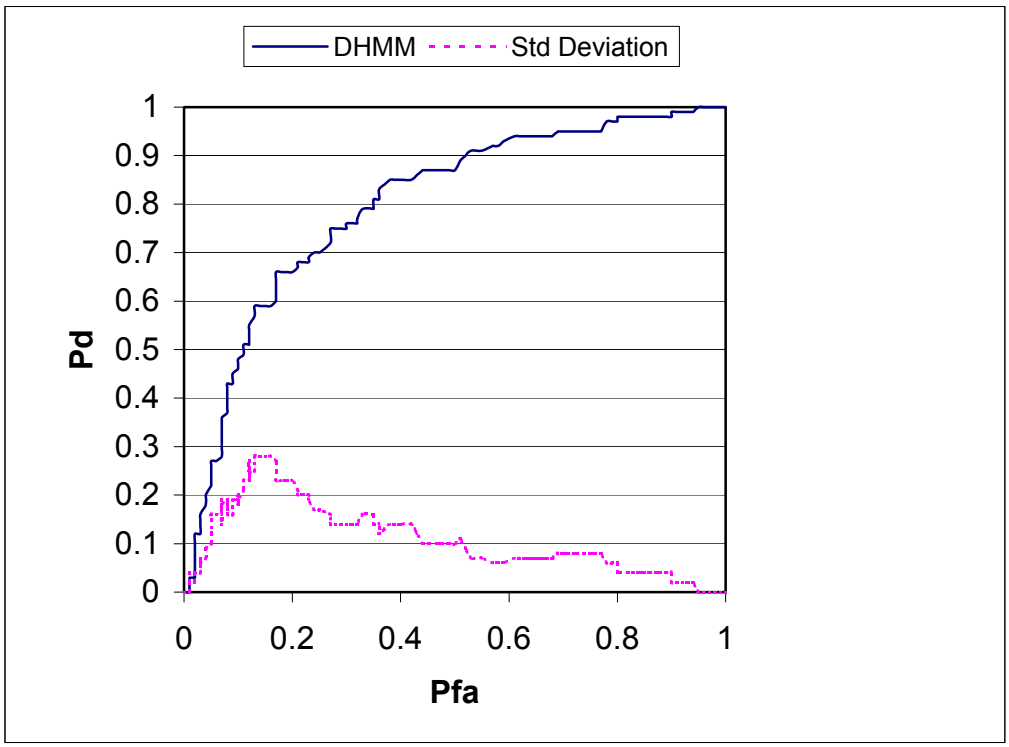

Figure 5: Results of Continuous HMM with 4 Gaussian Mixtures

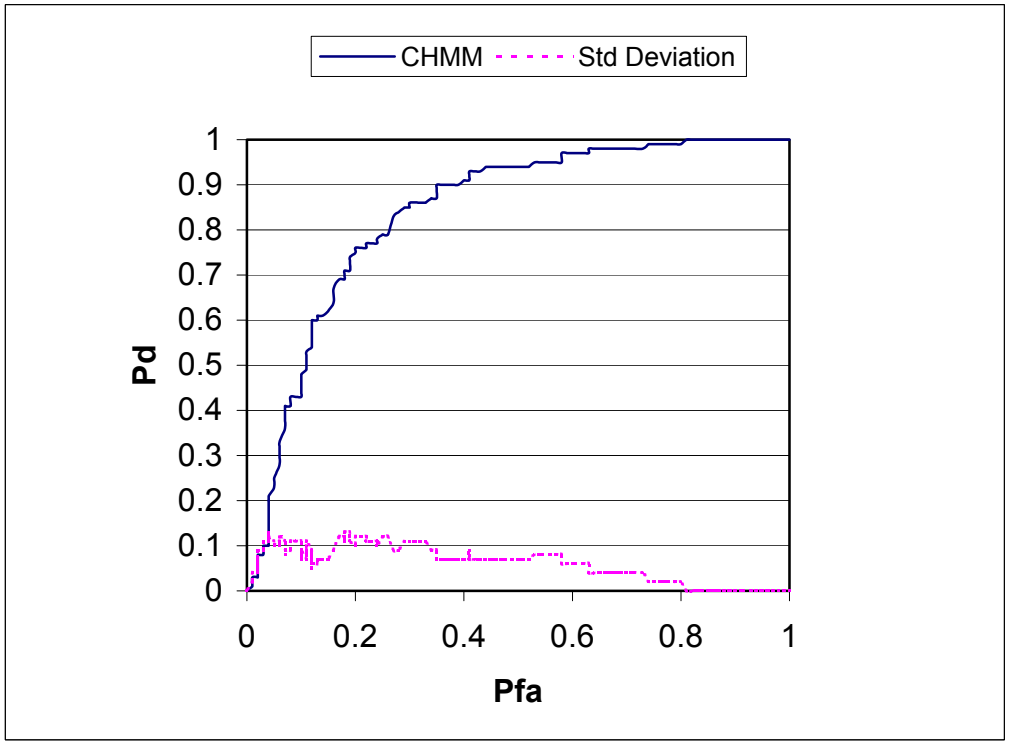

Figure 6: Results of Continuous HMM with 5 Gaussian Mixtures

4.2.2. DHMM Results

Two experiments were conducted with DHMM models. The first experiment used the train one/test three strategy. The second used the leave-one-out strategy. The DHMM models were initialized using the SOFM strategy described in [2], and then trained on one data set, further tested on other three data sets. We did this experiment with different codebook sizes. The results are shown in the figures 7-9.

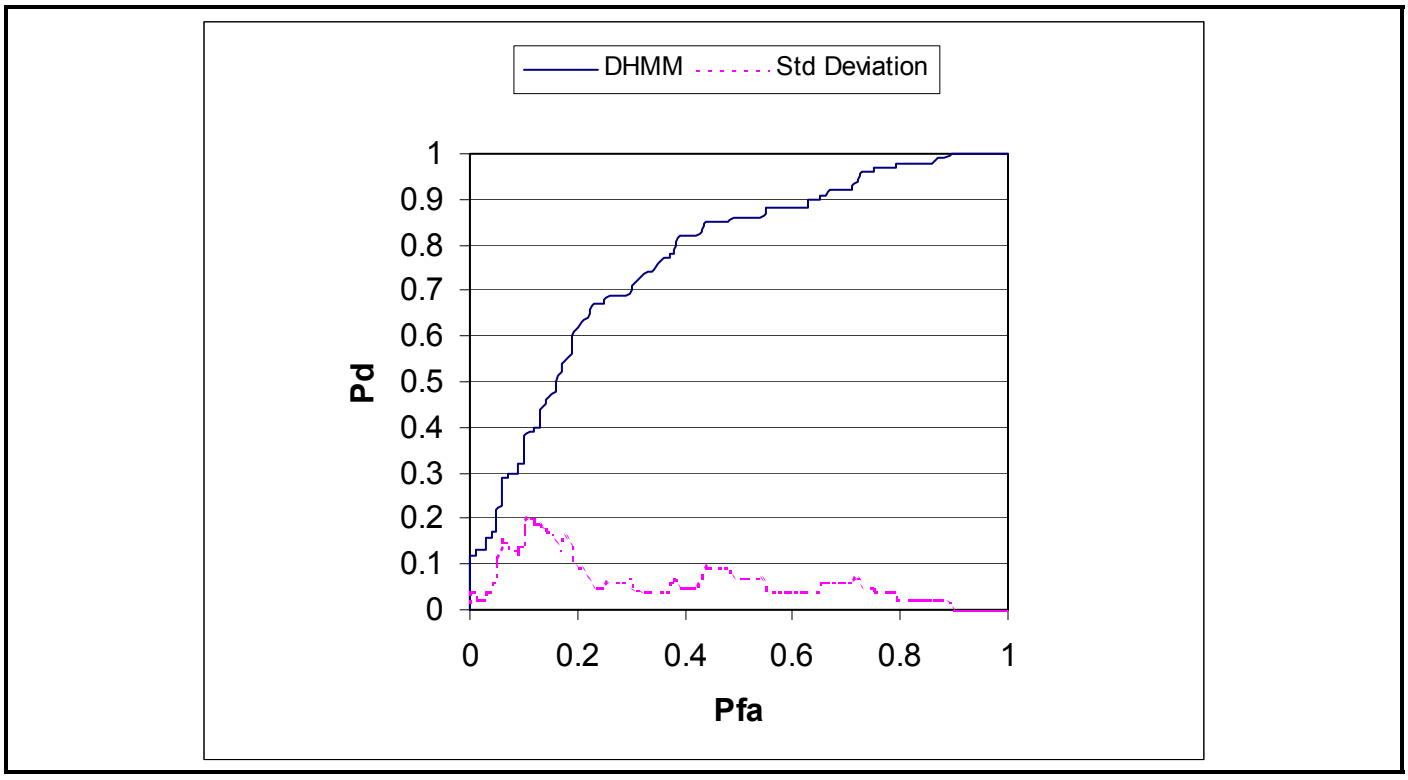

Figure7: Results of Discrete HMM with 25 Symbols in the Codebook using the train one/test three strategy.

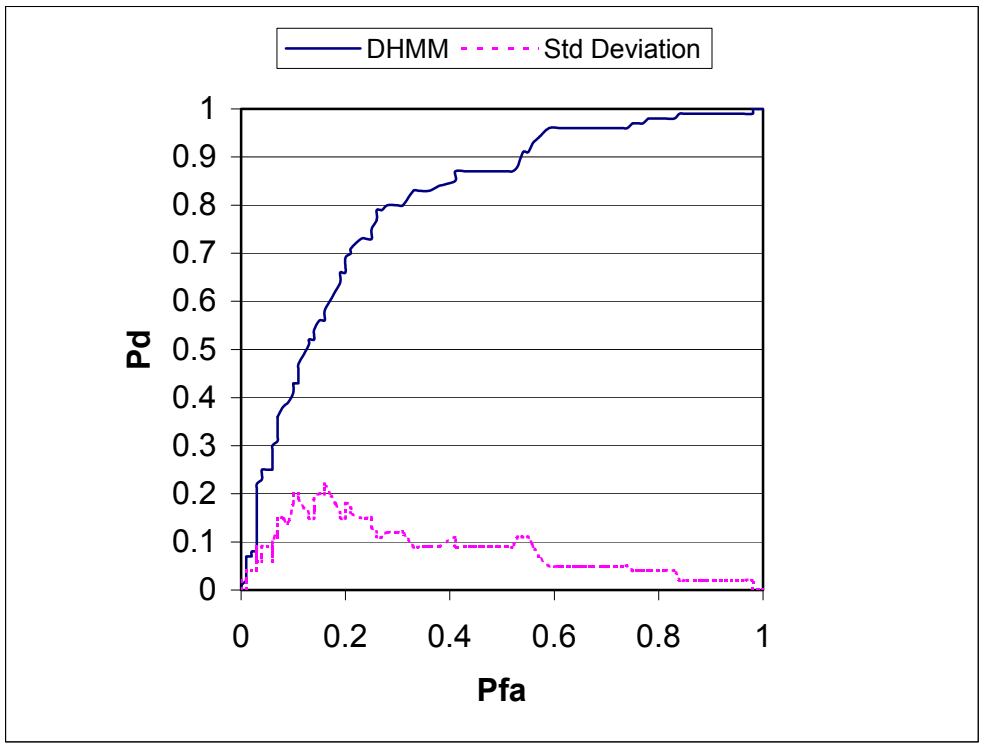

Figure 8: Results of Discrete HMM with 50 Symbols in the Codebook using the train one/test three strategy.

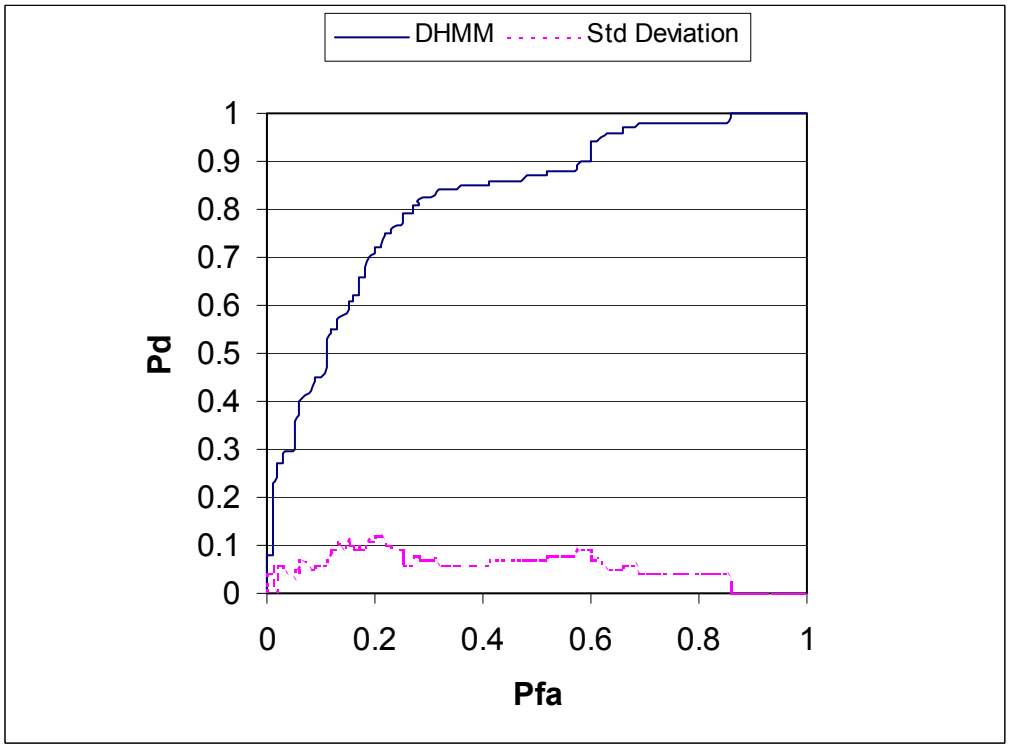

Figure 9: Results of Discrete HMM with 100 Symbols in the Codebook using the train one/test three strategy.

Discrete models were also trained with the leave-one-out strategy. The results are shown below.

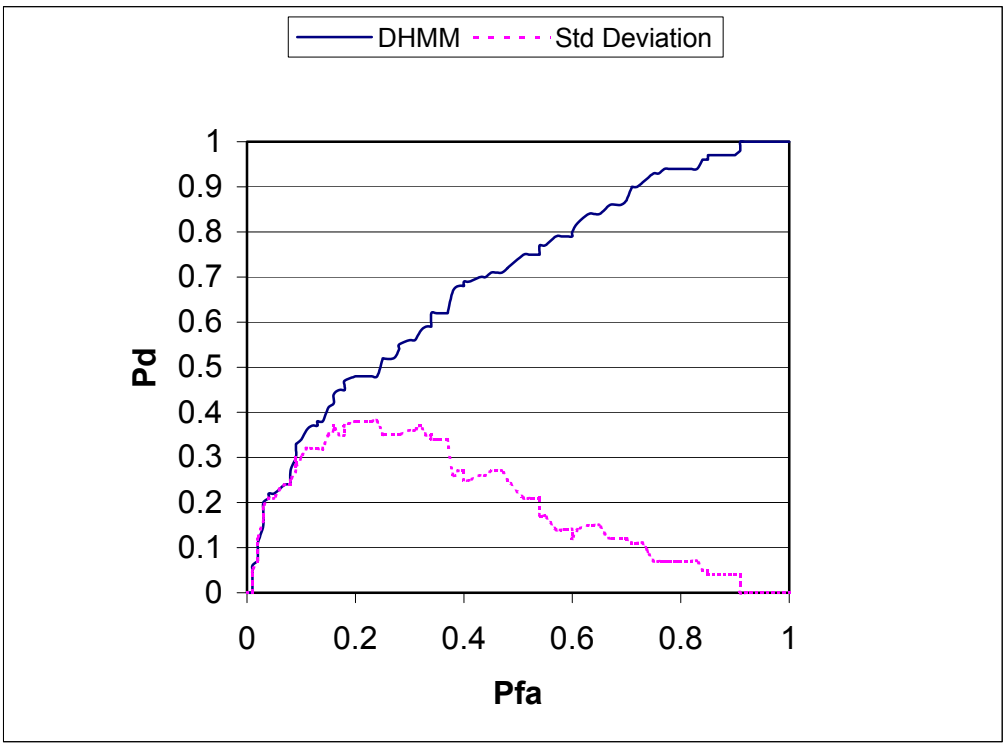

Figure10: Results of Discrete HMM with 25 Symbols in the Codebook using the leave-one-out strategy.

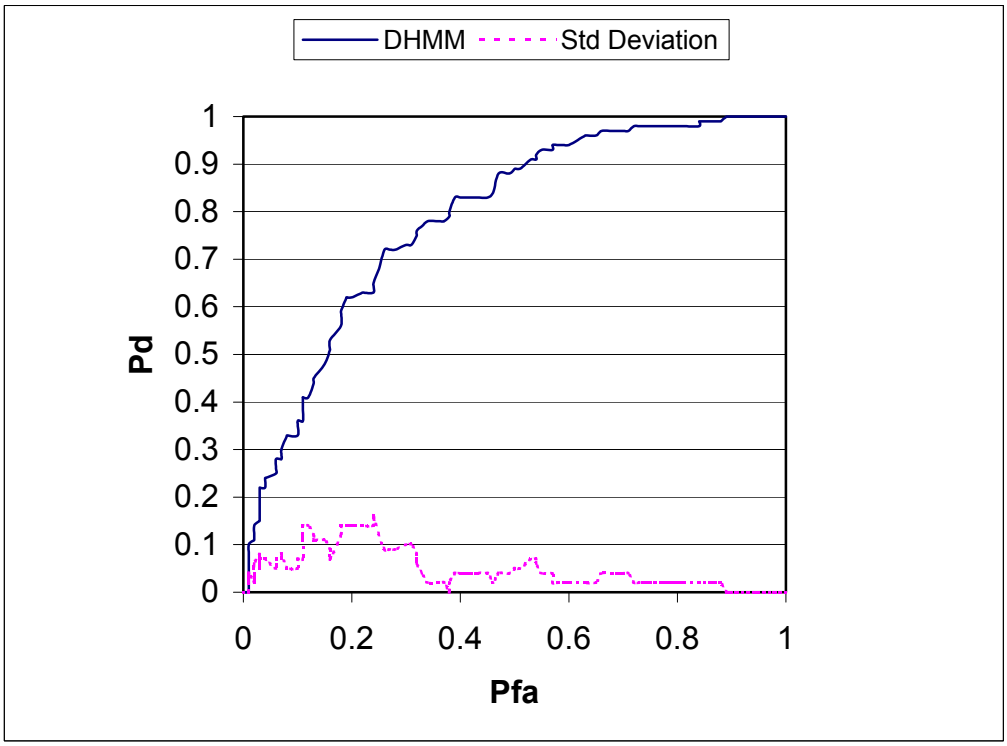

Figure 11: Results of Discrete HMM with 50 Symbols in the Codebook using the leave-one-out strategy.

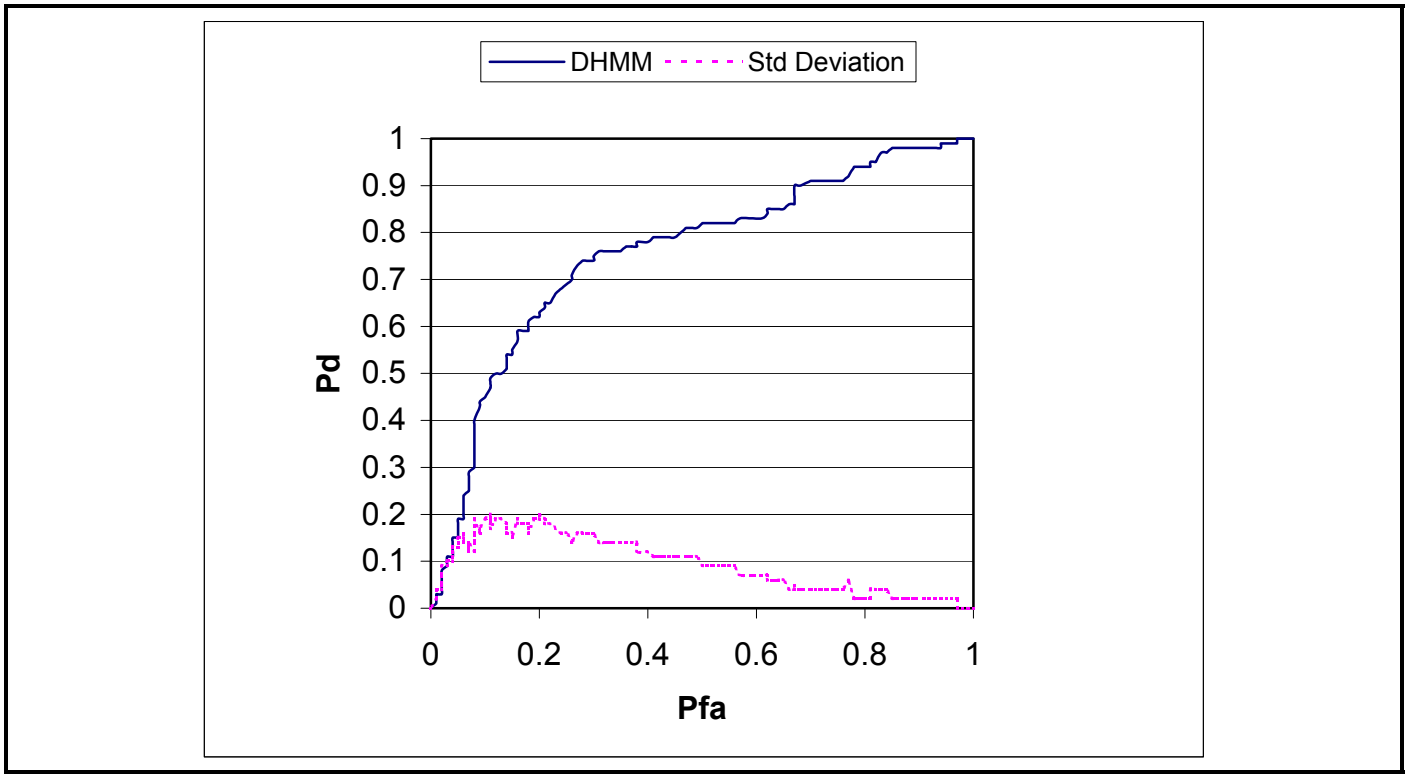

Figure12: Results of Discrete HMM with 100 Symbols in the Codebook using the leave-one-out strategy.

4.2.3. Fusion Results

We did a simple fusion to the outputs of CHMMs and DHMMs. The fusion was performed by averaging of the confidence values from each model for each grid square. The CHMM's results are shown in Figure 8. The DHMM's model is picked from the DHMM experiment associated with Figure 10. The fusion results are illustrated in Figure 13.

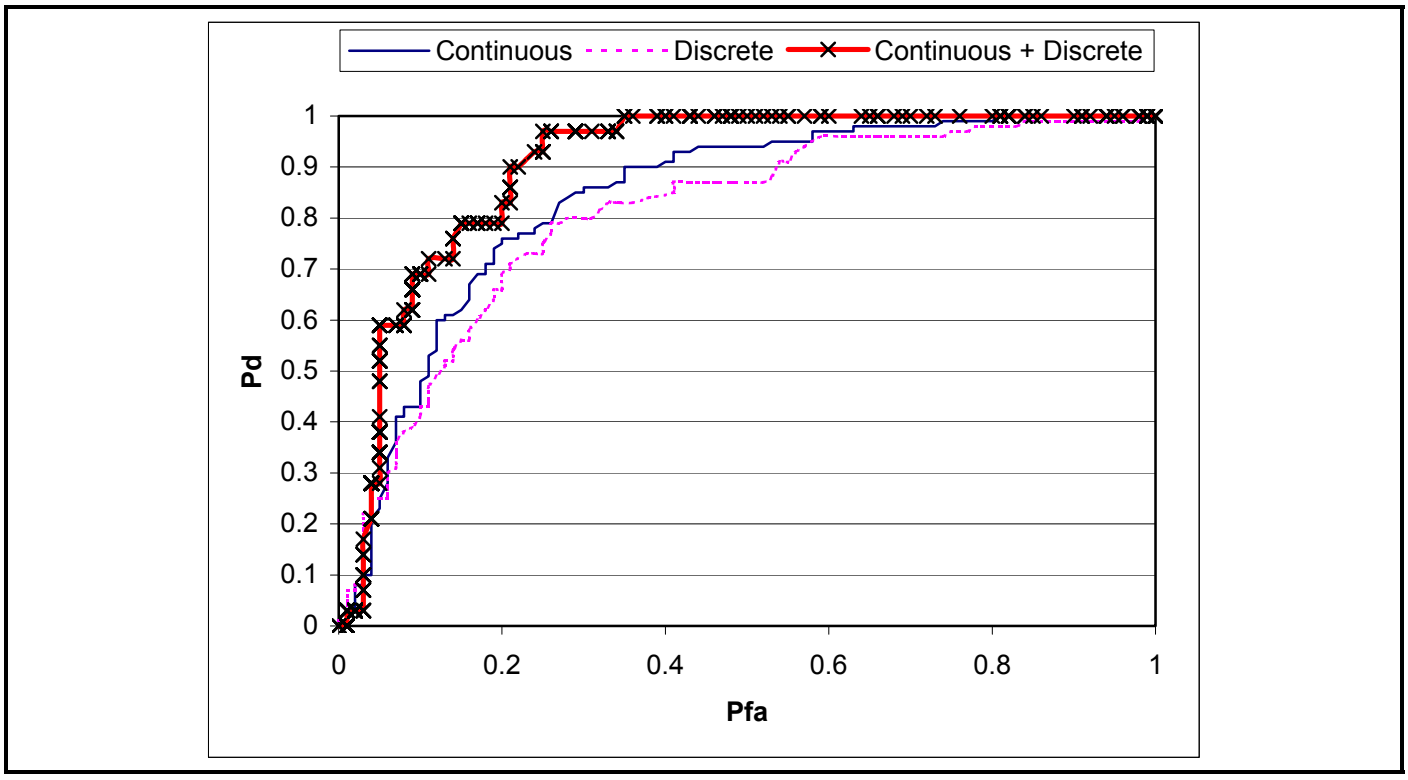

Figure 13: Results of Fusing the Continuous and the Discrete HMM Compared to Those of the Continuous and the Discrete HMM respectively

The fusion result performs better than that of each individual model. For example, at 100% and 90% PD, the PFA dropped by about ½ compared to the continuous model. And the continuous model performed a little better than the discrete model. These results can be compared to methods reported on in [10] generated by a hand-held GPR on the JUXOCO grid. The HMM with the EFGPR performs better than the baseline hand-held method that uses pointwise processing but not as well as the spatial processing method. However, there are several differences in experimental design that make the

comparison less than conclusive. For example, in [10] the positioning was much more precise, which is a significant issue in detection and discrimination of AP mines.

5. Summary

Continuous and discrete HMM algorithms were applied to GPR data acquired with the EFGPR system on the JUXOCO calibration grid. The algorithms were trained using Baum-Welch and MCE training. Two training strategies, the train one/test three and leave-one-out strategies were employed. Discriminative training demonstrates improved performance over Baum-Welch training. The best results were obtained with a fusion of continuous and discrete HMMs. The system achieved Probabilities of Detection of 100% and 90% with a Probabilities of False Alarm of about 40% and 25% respectively. These rates compare favorably with some published rates in the literature but not as well as the best. We conclude that HMMs with the existing feature sets perform well for detection of AT mines but discrimination between mines and discrete clutter objects requires better feature sets.

6. Bibliography

[1] Lawrence R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, vol. 77, pp. 257-285, Feb. 1989.

[2] P. D. Gader, M. Mystkowski, and Y. Zhao, "Landmine Detection with Ground Penetrating Radar Using Hidden Markov Models", *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, pp. 1231–1244, June 2001

[3] Y. Zhao, P. D. Gader, P. Chen, Y. Zhang, "Training DHMMs of mine and clutter to minimize landmine detection errors", *IEEE Trans. Geoscience and Remote Sensing*, vol. 41, No. 5, pp. 1016-1024,May 2003.

[4] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains", *The Annals of Mathematical Statistics*, vol. 41, pp.164-171, 1970

[5] L. A. Liporace, "Maximum Likelihood Estimation for Multivariate Observations of Markov Sources", *IEEE Trans. Information Theory*, vol. 28, pp. 729-734, Sep. 1982.

[6] B. H. Juang, S. Levinson, M. Sondhi, "Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chains", *IEEE Trans. Information Theory*, vol. 32, pp. 307-309, Mar. 1986.

[7] B. Juang, and S. Katagiri, "Discriminative Learning for Minimum Error Classification", *IEEE Trans. Signal Processing*, vol. 40, pp. 3043-3054, Dec 1992

[8] S. Amari, "A Theory of Adaptive Pattern Classifiers", *IEEE Trans. Electronic Computers,* vol. EC-16, pp. 299-307, Jun.1967

[9] L. H. Chen, J. R. Chen, D. Liang, S. H. Deng, H. Y. Liao, "A Minimum Classification Error Method for Face Recognition", *Image Processing and Its Applications (Seventh International Conference)*, vol. 2, pp. 630-633, Jul. 1999.

[10] K. C. Ho, L. M. Collins, L. G. Huettel, P. D. Gader, Discrimination Mode Processing for EMI and GPR sensors for Hand-Held Land Mine Detection, *IEEE Trans. Geoscience and Remote Sensing* (accepted).